

A Multi-view Classification Framework for Falls Prediction: Multiple-domain Assessments in Parkinson's Disease

Xiuyu Huang
The University of Sydney; Centre
for Smart Health, The Hong Kong
Polytechnic University
xhua0383@uni.sydney.edu.au

Mark Latt
The University of Sydney
mark.latt@sydney.edu.au

Matloob Khushi
The University of Sydney
matloob.khushi@sydney.edu.au

Paulo Pelicioni
Neuroscience Research Australia
p.pelicioni@neura.edu.au

Matthew Brodie
Neuroscience Research Australia
matthew.brodie@neura.edu.au

Stephen Lord
Neuroscience Research Australia
S.Lord@neura.edu.au

Clement Loy
The University of Sydney
clement.loy@sydney.edu.au

Simon K. Poon
The University of Sydney
simon.poon@sydney.edu.au

Abstract

Falls are one of the most common causes of injury and disability in people with Parkinson's disease (PD). This study developed an augmented machine learning framework for screening the risk of falling in people with PD using multiple domain assessments. A sample of 109 people with PD (50 fallers and 59 non-fallers) undertook four domains of assessment: disease-specific rating scales, clinical examination measures, physiological assessments, and gait analysis. A multi-view classifying framework was developed from a sequence of procedures and achieved 77.50% average predicting accuracy. The robustness of the multi-view framework was tested by comparing outcomes of three different view selection methods. The developed framework may have implications for clinical decision making, as some of the PD fall risk variables/features may be amenable to treatment. Our results showed that external reliability can be achieved by a simple voting mechanism from multiple, perhaps diverse, perspective consensus.

1. Introduction

Parkinson's disease (PD) is a common neurodegenerative disorder, with distinct clinical features including bradykinesia, hypometria and tremor [1]. Falling occurs frequently among people with PD [2]. A large portion of people with PD (38-68%) fall yearly [3, 4]. Most people (87%) who have had a

diagnosis of PD for more than 20 years have fallen at least once [5]. The consequences of falling include reduced mobility, injury, pain, loss of independence, stress and reduced quality of life. As falling in this population is an important health care challenge, it is important to develop models to predict falls and assess risk factors that may be amenable to treatment.

Machine learning (ML) is a great role in public health research [6]. It has been applied to tremor assessment [7], instantaneous fall detection [8] and identification of freezing of gait (FOG) [9] among people with PD. A large number of fall risk factors, including high total Unified Parkinson's Disease Rating Scale (UPDRS) score [10], advanced Hoehn and Yahr (HY) stage [11] and older age [12], have been investigated in many cohort, cross-sectional and case-control studies. Generally, these factors can be categorized as: disease-specific rating scales, clinical examination measures, physiological assessments and gait analysis [13]. A meta multi-view classifying model, an innovative ML tool [14, 15], may be useful in the analysis of variables from multiple categories. It aims to generate a learning function/classifier to model each view separately, optimize all functions in a final model and increase generalization performance. It combines the powerful computational ability of ML in handling high-dimensional data with learning from multiple perspectives.

We aimed to build a multi-view classifier to predict falls in people with PD. The remainder of this paper is organized as follows: Section 2 - Background; Section

3 - Methodology; Section 4 - Results; Section 5 - Discussion; and Section 6 - Conclusion.

2. Background

Identifying risk factors for falls may assist clinical interventions to prevent falls. Fall risk factors among people with PD can be categorized using different criteria. For example, in one study [16], risk factors were categorized as non-remediable (irreversible) or potentially remediable. Alternatively, risk factors may be divided into four categories: disease-specific rating scales, clinical examination measures, physiological assessments and gait analysis [13].

Among people with PD, fall prediction models may inform health care decisions [17]. In many hypothesis driven statistical studies [10, 18-23], univariate and multivariate logistic regressions were used, with sensitivities ranging from between 77% and 97% and specificities between 71% and 100%. Disease-specific rating scales identify or quantify symptoms, document disease progression and measure response to treatment. Many PD-specific scales [10, 11, 20, 24] are significantly associated with falls. Recent prospective studies have also identified significant associations between falls and clinical examination measures [12, 18, 25, 26], including demographic and PD-specific variables. Eleven physiological measures [13, 20, 21, 27, 28], such as leg strength and walking speed, may also contribute to fall risk. Gait analysis techniques, e.g. involving accelerometry [13, 29, 30], have been used to identify differences in gait pattern between fallers and non-fallers with PD. These studies examined the magnitude and rhythmicity of acceleration or extracted new features from the original accelerometry. Significant differences between fallers and non-fallers in PD have been found for those extracted features.

In assessing the current state of research, we identified three main areas for further exploration. First, many empirical studies we reviewed on fall prevention focused on one or two risk factor categories, not addressing the high-dimensional perspective [31]. Second, a variety machine learning approaches have been used in assisting PD diagnosis [14, 32-34], automatic tremor monitoring [7, 35], and FOG detection [36-38], but not in the area of fall prediction. Third, many machine learning based predictive models [10, 18-23] are yet to be evaluated on the validity as well as the reliability.

To address these issues, a multi-view classification framework was developed. Based on data domains, five different models were considered and configured through the majority voting strategy. Additionally, models were evaluated by a pipelined framework

(shown in the Methodology section) and found to predict fallers accurately.

3. Methodology

Our novel framework involved three phases: 1. 'data & preparation', 2. 'model development', and 3. 'model evaluation' (summarized in Figure 1 and expanded in Figure 2). In phase 1, 'data & preparation', data was (1) duplicated 20 times and (2) randomly split into 2 data subsets in each duplicated copy. Phase 2, 'model development', had 4 stages: (1) variable ranking and selection, (2) feature generation & feature selection, (3) hyperparameter tuning and (4) multi-view model configuration. Phase 3, 'model evaluation', consisted of two integrated steps: (1) unseen data prediction in each data copy and (2) merging results from all the duplicated copies.

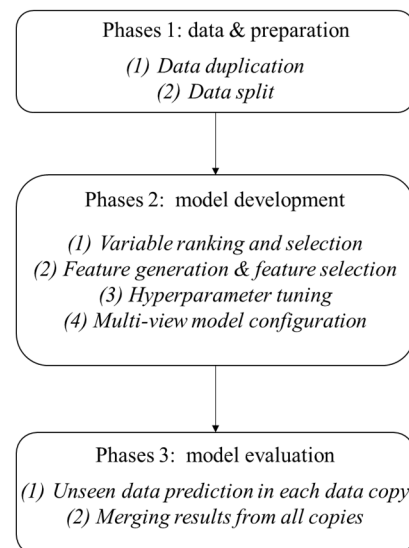


Figure 1. Overall multi-view classification framework

3.1. Phase 1: Data & preparation

A sample of 109 people with PD were recruited to participate in this study [13]. Participants were eligible if they had a diagnosis of idiopathic PD based on the United Kingdom PD Brain Bank criteria [39] and lived in the community. However, they were excluded if they were unable to walk independently without walking aids while doing the gait assessment or had psychotic symptoms or dementia. Participants were assessed when they experienced their usual improvement in mobility following anti-Parkinsonian medications (i.e. typical 'on' state), with assessments lasting up to 25 minutes. All participants were instructed to complete a Fall Diary each day and return completed sections each month for

one year and were also phoned monthly for one year to complete a structured interview regarding fall incidents. The protocol was approved by the Human Studies Ethics Committee at the University of Sydney and informed consent was obtained from all participants before their attendance.

The collected data was duplicated 20 times. In each duplicated copy, the data set was randomly split into two data subsets (N=87/109, 80%) for model development and (N=22/109, 20%) for testing. The model development data subset was used to establish the best ML models, and the testing data subset was used to assess the performance of the chosen ML models. Data copying aimed to eliminate random biases generated by data splitting and increase the reliability of model evaluation processes.

Domain 1: Disease-specific rating scale data included: Schwab and England Activities of Daily Living (SEADL) scale, HY scale, UPDRS) Mini-Mental State Scale (MMSE), Frontal Assessment Battery (FAB) and Abnormal Involuntary Movement Scale (AIMS).

Domain 2: Clinical examination data involved demographic data, clinical history and examination variables and PD-specific measures. Demographic data included gender, age, height and weight. Patients were asked about falls in the prior year, hip fractures, dizziness, angina or heart attack, urinary incontinence, faecal incontinence, number, class and dose of prescribed and consumed medications (such as levodopa preparations, dopamine agonists, other anti-parkinsonian medications and anticholinergic drugs)

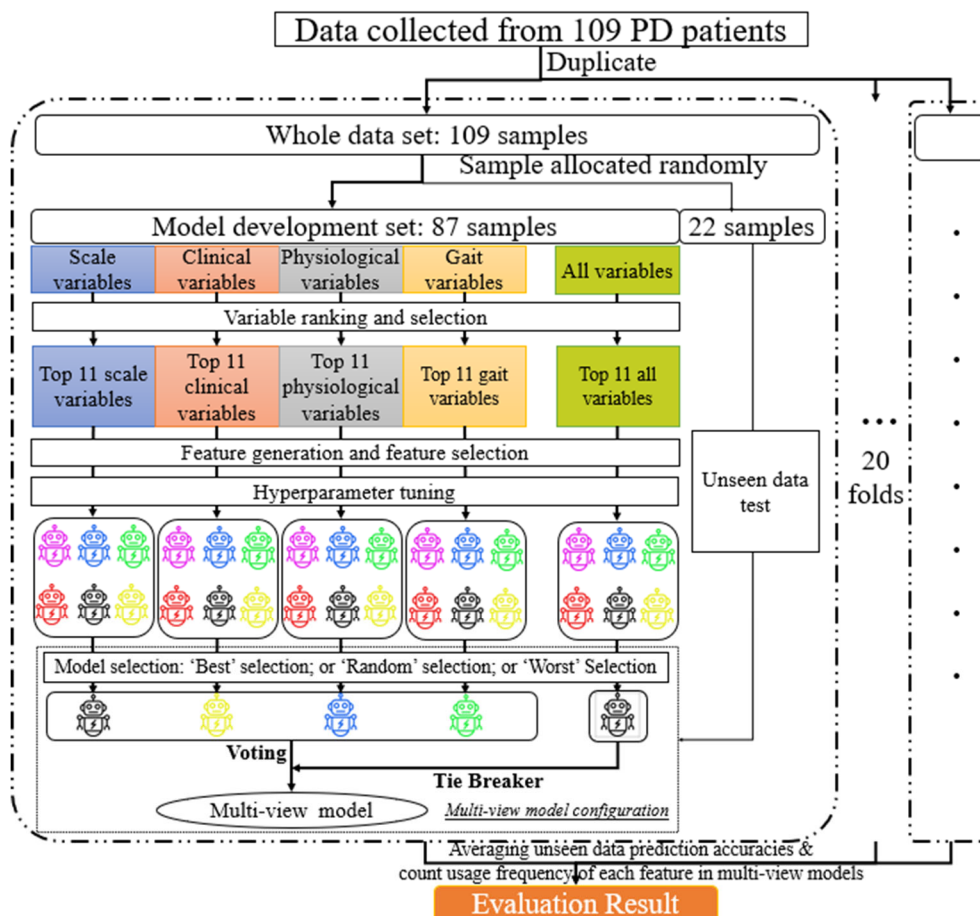


Figure 2. Overall structure of modeling and evaluating

* The models prior to 'model selection' had been optimized by feature selection and hyperparameter tuning.

* KNN, LR, DT, NB, RF, and SVM were tried in feature selectors, so 6 models were in each domain.

Four domains of data were collected from participants: disease-specific rating scales, clinical examination measures, physiological assessments and gait analysis. Collected variables were also concatenated as an aggregated domain dataset.

and time since the initial diagnosis of PD. Blood pressure (obtained using a mercury sphygmomanometer) and radial pulse were measured after lying supine for 10 minutes, instantly after rising to a standing position and after standing for three minutes. Limb rigidity, axial posture, bradykinesia, and

dyskinesia were assessed using part 3 of UPDRS (Motor Subscale).

Domain 3: Physiological assessment data was obtained using the Physiological Profile Approach (PPA) [40]: visual contrast sensitivity, low-contrast visual acuity, lower limb proprioception, tactile sensitivity, maximal isometric strength of muscle groups (measured via spring gauges), simple action time (measured by an electronic timer), postural sway (assessed with eyes open and closed, while standing on firm and compliant surfaces, using a sway meter attached to the waist), visual contrast sensitivity (Melbourne Edge Test), visual acuity (a standard LogMAR eye chart), low contrast visual acuity (LogMAR eye chart with grey lettering), proprioception error (difference in knee flexion angle between referent and test legs during position-matching tasks) and light touch sensation at the ankle (assessed using Semmes-Weinstein filaments).

Domain 4: Gait analysis data was collected using linear accelerometers [13]. A tri-axial piezo-electric accelerometer was firmly attached to the head of each participant via a light plastic frame, while another was attached posteriorly in the midline at the level of the sacrum/pelvis, through a Velcro belt. Linear accelerations at the head and pelvis were measured along three orthogonal axes: vertical (VT), anterior-posterior (AP) and medio-lateral (ML). Multiple variables were derived based on the approaches used in previous studies [41-43].

Domain 5: The aggregated domain was formed by combining all variables from the 4 different domains.

3.2. Phase 2: Model development

We followed a logical 4-step process in the model development phase: (1) variable ranking and selection, (2) feature generation and feature selection, (3) hyperparameter tuning and (4) multi-view model configuration. These processes were implemented using the model development subset of each duplicated data copy.

3.2.1. Variable ranking and selection. Univariate analyses were used for the ranking of variables. In univariate analysis, relationships between predictor variables and future falls were identified by t-tests for independent samples (continuous variables) and chi square tests for cross-tabulation (categorical variables). Significance levels were defined $p < 0.05$. The analyses were performed using SPSS Version 23 for Windows.

Based on p-values, variables were then ranked in order of decreasing significance. The top 11 variables were selected in each independent domain and in the aggregated domain for further model generation. In this study, we decided to keep the number of variables (as features) the same in each domain, hence highest number of variables was 11 collected in the scale domain.

Table 1. Number of collected variables

Domain	Number of variables
Scale	11
Clinical	23
Physiological	20
Gait	62
Aggregated	116

3.2.2. Feature generation and selection. Suitable variables were processed for exhaustive feature selection [44]. Categorical variables were encoded by label encoding as machine-learning features, and continuous variables were normalized by z-score as the input for further ML models.

An exhaustive feature selector was designed to determine the best feature subsets of the top 11 selected features in each domain for the classifying models. The selector was run using 10-fold cross-validation (10-fold CV) for six different commonly used ML classification models including K-Nearest Neighbors (KNN), Logistics Regression (LR), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). Exhaustive feature selection is a wrapper method which naively evaluates all combinations of feature subsets to find the best subset of features to build the best predictor.

3.2.3. Hyperparameter tuning. The above ML models were run with default hyperparameters. We improved their performance by tuning the hyperparameters via a grid search method. This method optimizes classifiers by exhaustively searching the best hyperparameters through a manually-defined subset of their hyperparameter spaces. The grid search algorithm was guided by 10-fold cross validation accuracy and was run for all classifiers that were defined in Section 3.2.2.

3.2.4. Multi-view model configuration. There were two steps in the configuration of multi-view models (shown in Figure 2). In the first step, we selected a suitable classifier according to one of the selection scenarios (explained in following paragraph) from the 6 classification models (KNN, LR, DT, NB, RF, SVM)

$$\text{Overall performance} = \frac{\sum_{i=1}^{20} P_i \text{ predicting accuracy of evaluation set (N=22,20\%) in each duplicated copy}}{20} \quad (1)$$

for the 5 domains (scale features, clinical features, physiological features, gait features and aggregated features).

The second voting step was to classify based on ‘a majority voting’ principle. Initially, we counted only the votes for each classifier for the predictions (0 for non-faller or 1 for faller) from 4 classifiers selected (one from each independent domain). The fifth classifier selected from the aggregated domain was used to break the tie when the predictions of two domains were different to the remaining two domains.

We tested the robustness of our voting framework by comparing 3 different scenarios (shown in Figure 2 and 3):

1) ‘Best Selection’, selecting the classifiers with the highest 10-fold cross-validation accuracy from each domain.

2) ‘Random Selection’, randomly selecting 1 out of 6 models from each domain.

3) ‘Worst Selection’, selecting the classifiers with the lowest 10-fold cross-validation accuracy from each domain.

3.3. Phase 3: Model evaluation

As shown in figure 2, the predictive accuracy of the evaluation set in each duplicated copy was recorded. As seen in the Equation (1), the overall performance of the multi-view model was represented by the average of predictive accuracies among all 20 duplicated copies. The performances of ‘Best’, ‘Random’ and ‘Worst’ selection multi-view models were calculated based on the equation. We also recorded frequencies of the used features in those multi-view models to obtain clinical insights into classifying models.

Table 2. Collected variables

Domain	Variables
Scale	The scores of SEADL, HY, UPDRS, MMSE, FAB, AIMS, and 5 subscales of UPDRS including Mentation/Behavior/Mood (MBE), Activities of Daily Living (ADL), Complication of Therapy (CoT), Other Complications (OC), and Motor Exam (ME).
Clinical	Gender, age, height, weight, Fall last year, Hip fracture, Dizziness, Angina or heart attack, Urinary incontinence, Faecal incontinence, Medications, Number of drugs, Blood pressure, Pulse (beats/min), Year since diagnosis, Leg/axial rigidity, Axial posture, Bradykinesia, Dyskinesia, FOG, Levodopa, Other dopamine agonist, Anticholinergic drug

Physiological Knee extension weaker leg^a (KEWLIS), knee extension stronger leg^a (KEWLIS), knee flexion weaker leg^a (KFWLIS), knee flexion stronger leg^a (KFSLIS), ankle dorsiflexion weaker leg^a (ADWLIS), ankle dorsiflexion stronger leg^a (ADSLIS), Hand Slower Side^b, Hand Faster Side^b, Foot Slower Side^b, Foot Faster Side^b, Contrast sensitivity score, High contrast (MAR), Low contrast (MAR), Proprioception, Light touch, Firm surface/eyes open^c, Firm surface/eyes closed^c, Compliant surface/eyes open^c, Compliant surface/eyes closed^c, score of Co-ordinated stability task

Gait Speed, Cadence, Step length, Step time variability, Head VT RMS^d, Head AP RMS^d, Head ML RMS^d, Head VT SD^e, Head AP SD^e, Head ML SD^e, Head VT HR^f, Head AP HR^f, Head ML HR^f, Head VT 8HR^g, Head AP 8HR^g, Head ML 8HR^g, Head VT Jerk^h, Head AP Jerk^h, Head ML Jerk^h, Head Jerk AP/VTⁱ, Head Jerk ML/VTⁱ, Head AP Vel^k, Head ML Vel^k, Head Vel Ramble^l, Head AP Ramble^m, Head ML Ramble^m, Head Ramble Areaⁿ, Head VT Range^o, Head AP Range^o, Head ML Range^o, Head VT VelRange^p, Head AP VelRange^p, Head ML VelRange^p,

Pelvis VT RMS^d, Pelvis AP RMS^d, Pelvis ML RMS^d, Pelvis VT SD^e, Pelvis AP SD^e, Pelvis ML SD^e, Pelvis VT HR^f, Pelvis AP HR^f, Pelvis ML HR^f, Pelvis VT 8HR^g, Pelvis AP 8HR^g, Pelvis ML 8HR^g, Pelvis VT Jerk^h, Pelvis AP Jerk^h, Pelvis ML Jerk^h, Pelvis Jerk AP/VTⁱ, Pelvis Jerk ML/VTⁱ, Pelvis AP Vel^k, Pelvis ML Vel^k, Pelvis Vel Ramble^l, Pelvis AP Ramble^m, Pelvis ML Ramble^m, Pelvis Ramble Areaⁿ, Pelvis VT Range^o, Pelvis AP Range^o, Pelvis ML Range^o, Pelvis VT VelRange^p, Pelvis AP VelRange^p, Pelvis ML VelRange^p,

^aIsometric strength (kg), ^bSimple action time (ms), ^cPostural sway (mm), ^dRoot mean square, ^eStandard deviation, ^fharmonic ratios, ^g8-step harmonic ratios, ^hroot mean square of the derivative of acceleration in each time stamp, ⁱRatio [43] of lateral to vertical AP/VT RMS^d Jerk^h [41], ^jRatio [43] of lateral to vertical AP/VT RMS^d Jerk^h, ^kVelocity of sway [41], ^lVelocity of the low frequency component of sway in the whole transverse plane (AP and ML combined), ^mLow frequency component of sway, ⁿLow frequency component of sway in AP axis multiple with low frequency component of sway in ML axis, ^o95% range of sway, ^p95% range velocity of sway

4. Result

4.1. Dataset

A total of 11 scale-domain, 23 clinical-domain, 20 physiological-domain, and 62 gait-domain variables were collected during the data collection process (Table 2).

All participants finished the one-year follow-up, and 50 participants (46%) reported that they fell one or more times during the year. There were 2159 falls in total, and 80% occurred during walking, mostly as a result of FOG. Of the falls, 138 lead to injury, with 3 hip fractures, 1 tibial fracture, and 1 radial fracture.

4.2. Model development measures

The classifying models of each domain were selected for configuring multi-view model in section

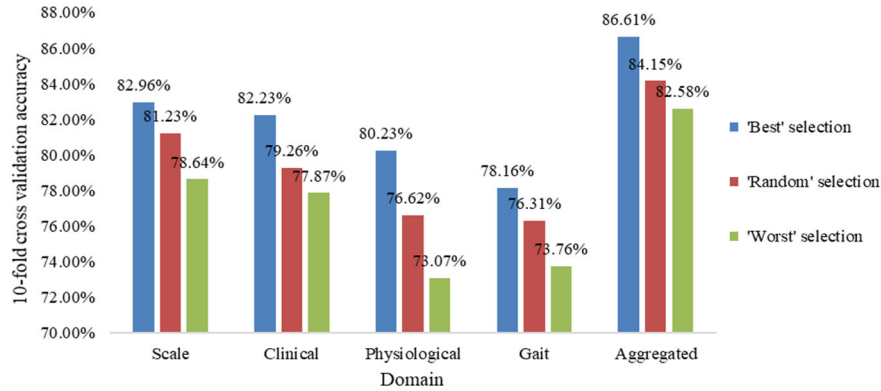


Figure 3. Comparison of average 10-fold validation accuracy of single view in model development process

3.2.4, after model optimizations (including feature selection and hyperparameter tuning) were implemented. The average 10-fold validation accuracy of single view (i.e. average of all the duplicated copies) of three different model selection methods is presented in Figure 3. The average 10-fold validation accuracies of the ‘Best’ model in each view were all higher than 78%, and the highest accuracy achieved was 86.61% (single aggregated domain). The average 10-fold validation accuracies of the ‘Random’ model in each view were all above 76%, and the highest (aggregated domain) achieved 84.15%. The average 10-fold validation accuracies of the ‘Worst’ model in each view were all above 73%, and the highest (aggregated domain) achieved 82.58%. Single view accuracies achieved by aggregated domain (86.61%, 84.15%, and 82.58% respectively) performed better than other single domains in all three scenarios.

4.3. Model evaluation

Our results were based on the prediction accuracies assessed by the 20 duplicated copies of the unseen 20% holdout data subsets. As shown in table 3, the average predictive accuracies achieved by the “best” selection, “random” selection, and “worst” selection scenario in single view were respectively between 53.85% and 70.91%, between 55.68% and 74.32 %, and between 56.36% and 74.32 %. By comparing these results to those in Figure 3, we found that predictive accuracies from holdout (testing) data subsets were lower than those from model development (training) data subsets. We also identified a pattern in model development and evaluation, i.e. results from the aggregated domain were always better than those from other single domains.

Average predictive accuracies achieved in the ‘best’, ‘random’ and ‘worst’ selection scenarios by multi-view were 75.68%, 76.14% and 77.50% respectively. Multi-view models performed better than single view ones in all three scenarios. More

interestingly, the performances of multi-view models from “worst” (77.50%) and ‘random’ (76.14%) scenarios were higher than all the single view models (between 53.85% and 70.91%) as well as the multi-view model (75.68%) in the “best” scenario.

Table 3. Comparison of average accuracy of each view and multi view model

View	Model Selection scenario		
	‘Best’	‘Random’	‘Worst’
Scale	67.95%	72.50%	74.32%
Clinical	66.82%	67.05%	70.91%
Physiological	65.91%	60.00%	65.68%
Gait	53.86%	55.68%	56.36%
Aggregated	70.91%	74.32%	72.27%
Multi-view	75.68%	76.14%	77.50%

Table 4. Two most frequently used features in the multi-view classifying models in all copies

Domain	Frequency rank	
	1 st	2 nd
Scale	HY: 57	ADL: 54
Clinical	Fall last year: 59	Axial posture: 57
Physiological	KEWLIS:50	score of Co-ordinated stability task: 48
Gait	Pelvis Jerk AP/VT: 55	Head Ramble Area: 53
Aggregated	HY: 55	Fall last year: 50

* Maximum number of frequencies = 3 (number of multi-view models in a data copy) * 20 (number of copies) = 60

As shown in Table 4, the two most frequent scale domain features in the multi-view classifying models were HY (57 times) and ADL (54 times). In the clinical domain, fall last year (59 times) was the most frequent feature, followed by Axial posture (57 times). In the

physiological domain, KWLIS (50 times) and the score of Co-ordinated stability task (48 times) were the most frequent in the multi-view models in all the data copies. Pelvis Jerk AP/ML (55 times) and Head Ramble Area (53 times) were the most frequent in the multi-view classifying model in the gait domain. In the aggregated domain, HY (55 times) and fall last year (50 times) were the two most frequent features.

5. Discussion

5.1. Clinical explanation and interpretation

Regarding the scale domain, HY stage and UPDRS ADL subscale score were used most frequently in developing multi view classifying models. HY is a standard staging system for PD severity, with higher stages (i.e. worse postural stability in the present) leading to higher incidence of future falls among people with PD. Greater impairments in activities of daily living (i.e. higher UPDRS ADL subscale scores) increased the incidence of future falls, in line with previous research [45].

In the clinical domain, a history of a fall in the prior year and abnormal axial posture were the two most important features. A history of falling is a recognized risk factor for future falls among people with PD and older people but has only limited predictive power on its own. A meta-analysis of prospective studies of falls in PD found that only 57% of participants who fell during a 3-month follow up [20]. Our study demonstrated that a history of a fall has a powerful predictive ability when aggregated with other variables/features. Abnormal axial posture decreases limits of stability [46], thus contributing to future falls.

In physiological data, isometric knee extension strength in the weaker leg and coordinated stability score are significant predictors of falls. Muscle strength has been identified as a significant fall risk for older people [40]. Isometric knee extension strength in the weaker leg was used as an entry feature for all models in this category. Our results confirmed that coordinated stability test score is a significant predictor for future falls in PD. It was used to quantify subjects' ability to adjust their balance near to their limits of stability during a mobility (trail drawing) task. This type of postural stability is crucial because the majority of PD fall events happen as a result of balance control loss in daily life [2].

For gait data, the Pelvis Jerk AP/VT and Head Ramble Area played a key role. This may suggest that lower sagittal- and medio-lateral pendulum stability may contribute to future falls in PD.

In the aggregated domain, no single domain features (top 3 were HY, fall last year, and KEWLIS) can dominate one classifier for predicting future falls in PD. This may indicate that each of the domains is complementary to each other in increasing fall risk.

5.2. Prediction reliability and validity of the multi-view model

In contrast to the mainstream ML approaches that aim to optimize fitness (or minimize errors) throughout the development process from feature selection, hyperparameter tuning and algorithm selection, we have shown our multi-view framework that it is possible to achieve reliable prediction performance without optimizing every step along the ML generation pipeline. Our approach uses consensus based on assessment criteria from their respective domains, hence enabling us to reduce the biases due to overfitting from a particular domain with even higher predictive performance from the training phase. Our findings shown in Table 3 demonstrated that it is not necessary to select the best model in each domain to achieve best performance, on the contrary, a multi-view model using the worst ML model in each domain could achieve the similar predictive performance when best individual ML model were selected. This interesting phenomenon may due to the fact that algorithms generated by the data in each domain all had an accurate fall prediction to specific PD patients, and our designed multi-view pipeline inherited their advantages through consensus mechanism no matter whether they are the best algorithms within its domain or not. It suggests that not only it is not critical to search for the best predictive model in each domain, and also using the multiple perspective voting mechanism could potentially provide a more robust and reliable prediction. Unlike the ensemble approach, the final classification relies on the consensus of majority (perhaps the collective wisdom) from each individual (or even weaker) classification model. This super-additive (or complementarity) effect produced in the multi-view rating process could play a critical role in decision making when dealing with imperfect ML models.

5.3. Limitations

One of the limitations could have affected our results was the participant inclusion criteria. Data were collected only during an 'on phase' period; this may have reduced the sensitivity of the measures, as subjects in an 'on state' normally have a better performance than in an 'off state' in mobility and balance tests. Further

studies can address this limitation by collecting data in 'off' conditions or repeating measures over time, in both 'on' and 'off' states. Another limitation was the small sample size in relation the large number of variables measured and collected from each individual participant. This can be address by recruitment of participants.

6. Conclusion

This study used an integrated stepwise framework in the development of a model to differentiate fallers from non-fallers with PD using multiple domains – i.e. disease-specific rating scales, clinical examination measures, physiological assessments, and gait analysis. The variables that were highly associated with future falls were included as features for ML classifiers. Using exhaustive feature and hyperparameter selectors, we optimized the six common classification models (KNN, LR, DT, NB, RF, and SWM) that could differentiate optimally between fallers and non-fallers in each domain as well as in an aggregated domain. We validated our findings by using three view ('Best', 'Random' and 'Worst') scenarios from these models and built three multi-view classifying models based on a 'majority of vote' method.

Results demonstrated that the multi-view classifying model including variables from multiple domains could be a reliable tool to differentiate between

fallers and non-fallers in PD. By comparisons between 1) multi-view and individual view, 2) multi-view and aggregated domain view, and 3) three model selections approaches, we found that the consensus of multi-view offered a more objective and robust mechanism to optimize predictive accuracy and measurement validity. As many of the used PD fall risk variables/features are amenable for fall prevention interventions, the developed model may also have implications for clinical management.

To address the limitations of our findings and increase the reliability of our model, further developments could be done in these three ways: 1) collect more data in 'off' state conditions, 2) recruiting more participants, and 3) expand more candidate domains for the current multi-view model.

The multi-view approach can be easily expanded by including additional domains. Different ML algorithm can be used based on the suitability of the domain. There will be circumstances that we will need to consider using Blackbox ML models. For example, certain domains in image or text analysis from clinical records, we could have the flexibility of choosing a mix models including the ones that are not interpretation nor explainable. The next step would be to explore if there is a minimum predictive performance of the ML model from each domain to qualify the participation in the multi-view approach.

6. Reference

1. de Rijk, M.C., et al., *Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts*. Neurology, 2000. **54**(11): p. S21-S23.
2. Bloem, B.R., *Postural instability in Parkinson's disease*. Clinical Neurology and Neurosurgery, 1992. **94**: p. S41-S45.
3. Schrag, A., Y. Ben-Shlomo, and N. Quinn, *How common are complications of Parkinson's disease?* Journal of Neurology, 2002. **249**(4): p. 419-423.
4. Wood, B.H., et al., *Incidence and prediction of falls in Parkinson's disease: a prospective multidisciplinary study*. Journal of Neurology Neurosurgery and Psychiatry, 2002. **72**(6): p. 721-725.
5. Hely, M.A., et al., *The Sydney multicenter study of Parkinson's disease: The inevitability of dementia at 20 years*. Movement Disorders, 2008. **23**(6): p. 837-844.
6. Giger, M.L., *Machine Learning in Medical Imaging*. Journal of the American College of Radiology, 2018. **15**(3): p. 512-520.
7. Roy, S.H., et al., *High-resolution tracking of motor disorders in Parkinson's disease during unconstrained activity*. Movement Disorders, 2013. **28**(8): p. 1080-1087.
8. Yu, S., et al. *Motion sensor-based assessment on fall risk and Parkinson's disease severity: a deep multi-source multi-task learning (DMML) approach*. in 2018 IEEE International Conference on Healthcare Informatics (ICHI). 2018. IEEE.
9. Silva de Lima, A.L., et al., *Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review*. Journal of Neurology, 2017. **264**(8): p. 1642-1654.
10. Voss, T.S., et al., *Fall frequency and risk assessment in early Parkinson's disease*. Parkinsonism & Related Disorders, 2012. **18**(7): p. 837-841.
11. Gray, P. and K. Hildebrand, *Fall risk factors in Parkinson's disease*. The Journal of neuroscience nursing : journal of the American Association of Neuroscience Nurses, 2000. **32**(4): p. 222-8.
12. Koller, W.C., et al., *Falls and Parkinson's disease*. Clinical Neuropharmacology, 1989. **12**(2): p. 98-105.
13. Latt, M.D., et al., *Acceleration patterns of the head and pelvis during gait in older people with parkinson's disease: a comparison of fallers and nonfallers*. Journals of Gerontology Series a-Biological Sciences and Medical Sciences, 2009. **64**(6): p. 700-706.
14. Caramia, C., et al., *IMU-based classification of Parkinson's disease from Gait: a sensitivity analysis on sensor location and feature selection*. Ieee Journal of Biomedical and Health Informatics, 2018. **22**(6): p. 1765-1774.

15. Xu, C., D. Tao, and C. Xu, *A survey on multi-view learning*. arXiv preprint arXiv:1304.5634, 2013.
16. Canning, C.G., S.S. Paul, and A. Nieuwboer, *Prevention of falls in Parkinson's disease: a review of fall risk factors and the role of physical interventions*. Neurodegenerative disease management, 2014. **4**(3): p. 203-21.
17. Laupacis, A., N. Sekar, and I.G. Stiell, *Clinical prediction rules - a review and suggested modifications of methodological standards*. Jama-Journal of the American Medical Association, 1997. **277**(6): p. 488-494.
18. Ashburn, A., et al., *A community-dwelling sample of people with Parkinson's disease: characteristics of fallers and non-fallers*. Age and Ageing, 2001. **30**(1): p. 47-52.
19. Custodio, N., et al., *Predictive model for falling in Parkinson disease patients*. eNeurologicalSci, 2016. **5**: p. 20-24.
20. Kerr, G.K., et al., *Predictors of future falls in Parkinson disease*. Neurology, 2010. **75**(2): p. 116-124.
21. Matinolli, M., et al., *Recurrent falls and mortality in Parkinson's disease: a prospective two-year follow-up study*. Acta Neurologica Scandinavica, 2011. **123**(3): p. 193-200.
22. Paul, S.S., et al., *Three simple clinical tests to accurately predict falls in people with Parkinson's disease*. Movement Disorders, 2013. **28**(5): p. 655-662.
23. Pickering, R.M., et al., *A meta-analysis of six prospective studies of falling in Parkinson's disease*. Movement Disorders, 2007. **22**(13): p. 1892-1900.
24. Kataoka, H., et al., *Risk of Falling in Parkinson's Disease at the Hoehn-Yahr Stage III*. European Neurology, 2011. **66**(5): p. 298-304.
25. Balash, Y., et al., *Falls in outpatients with Parkinson's disease - frequency, impact and identifying factors*. Journal of Neurology, 2005. **252**(11): p. 1310-1315.
26. Mak, M.K.Y. and M.Y.C. Pang, *Fear of falling is independently associated with recurrent falls in patients with Parkinson's disease: a 1-year prospective study*. Journal of Neurology, 2009. **256**(10): p. 1689-1695.
27. Allen, N.E., et al., *Reduced muscle power is associated with slower walking velocity and falls in people with Parkinson's disease*. Parkinsonism & Related Disorders, 2010. **16**(4): p. 261-264.
28. Cole, M.H., et al., *Falls in Parkinson's Disease: kinematic evidence for impaired head and trunk control*. Movement Disorders, 2010. **25**(14): p. 2369-2378.
29. Hausdorff, J.M., *Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling*. Chaos, 2009. **19**(2).
30. Schaafsma, J.D., et al., *Gait dynamics in Parkinson's disease: relationship to Parkinsonian features, falls and response to levodopa*. Journal of the Neurological Sciences, 2003. **212**(1-2): p. 47-53.
31. Johnstone, I.M. and D.M. Titterton, *Statistical challenges of high-dimensional data introduction*. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 2009. **367**(1906): p. 4237-4253.
32. Cook, D.J., M. Schmitter-Edgecombe, and P. Dawadi, *Analyzing activity behavior and movement in a naturalistic environment using smart home techniques*. Ieee Journal of Biomedical and Health Informatics, 2015. **19**(6): p. 1882-1892.
33. Daliri, M.R., *Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease*. Biomedical Signal Processing and Control, 2013. **8**(1): p. 66-70.
34. Oung, Q.W., et al., *Empirical wavelet transform based features for classification of Parkinson's disease severity*. Journal of Medical Systems, 2018. **42**(2).
35. Patel, S., et al., *Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors*. Ieee Transactions on Information Technology in Biomedicine, 2009. **13**(6): p. 864-873.
36. Ahlrichs, C., et al., *Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients*. Medical & Biological Engineering & Computing, 2016. **54**(1): p. 223-233.
37. Kubota, K.J., J.A. Chen, and M.A. Little, *Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures*. Movement Disorders, 2016. **31**(9): p. 1314-1326.
38. Rodriguez-Martin, D., et al., *Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer*. Plos One, 2017. **12**(2).
39. Hughes, A.J., et al., *What features improve the accuracy of clinical diagnosis in Parkinson's disease: a clinicopathologic study*. Neurology, 2001. **57**(10): p. S34-S38.
40. Lord, S.R., R.D. Clark, and I.W. Webster, *Postural stability and associated physiological factors in a population of aged persons*. Journals of Gerontology, 1991. **46**(3): p. M69-M76.
41. Brodie, M.A.D., et al., *Head and pelvis stride-to-stride oscillations in gait: validation and interpretation of measurements from wearable accelerometers*. Physiological Measurement, 2015. **36**(5): p. 857-872.
42. Brodie, M.A.D., et al., *Uncontrolled head oscillations in people with Parkinson's disease may reflect an inability to respond to perturbations while walking*. Physiological Measurement, 2015. **36**(5): p. 873-881.
43. Menz, H.B., S.R. Lord, and R.C. Fitzpatrick, *Age-related differences in walking stability*. Age and Ageing, 2003. **32**(2): p. 137-142.
44. Hall, M.A. and L.A. Smith, *Feature subset selection: A correlation based filter approach*. Progress in Connectionist-Based Information Systems, Vols 1 and 2, ed. N. Kasabov, et al. 1998. 855-858.
45. Chang, J.T. and D.A. Ganz, *Quality indicators for falls and mobility problems in vulnerable elders*. Journal of the American Geriatrics Society, 2007. **55**: p. S327-S334.
46. Schieppati, M., et al., *The limits of equilibrium in young and elderly normal subjects and in parkinsonians*. Electroencephalography and Clinical Neurophysiology, 1994. **93**(4): p. 286-298.